The catalytic domain of avian sarcoma virus integrase: conformation of the active-site residues in the presence of divalent cations

Grzegorz Bujacz[1], Mariusz Jaskólski[1,2], Jerry Alexandratos[1], Alexander Wlodawer[1]*, George Merkel[3], Richard A Katz[3] and Anna Marie Skalka[3]

Background: Members of the structurally-related superfamily of enzymes that includes RNase H, RuvC resolvase, MuA transposase, and retroviral integrase require divalent cations for enzymatic activity. So far, cation positions are reported in the X-ray crystal structures of only two of these proteins, *E. coli* and human immunodeficiency virus 1 (HIV-1) RNase H. Details of the placement of metal ions in the active site of retroviral integrases are necessary for the understanding of the catalytic mechanism of these enzymes.

Results: The structure of the enzymatically active catalytic domain (residues 52–207) of avian sarcoma virus integrase (ASV IN) has been solved in the presence of divalent cations (Mn^{2+} or Mg^{2+}), at 1.7–2.2 Å resolution. A single ion of either type interacts with the carboxylate groups of the active site aspartates and uses four water molecules to complete its octahedral coordination. The placement of the aspartate side chains and metal ions is very similar to that observed in the RNase H members of this superfamily; however, the conformation of the catalytic aspartates in the active site of ASV IN differs significantly from that reported for the analogous residues in HIV-1 IN.

Conclusions: Binding of the required metal ions does not lead to significant structural modifications in the active site of the catalytic domain of ASV IN. This indicates that at least one metal-binding site is preformed in the structure, and suggests that the observed constellation of the acidic residues represents a catalytically competent active site. Only a single divalent cation was observed even at extremely high concentrations of the metals. We conclude that either only one metal ion is needed for catalysis, or that a second metal-binding site can only exist in the presence of substrate and/or other domains of the protein. The unexpected differences between the active sites of ASV IN and HIV-1 IN remain unexplained; they may reflect the effects of crystal contacts on the active site of HIV-1 IN, or a tendency for structural polymorphism.

Addresses: [1]Macromolecular Structure Laboratory, NCI-Frederick Cancer Research and Development Center, ABL-Basic Research Program, Frederick, MD 21702, USA, [2]Center for Biocrystallographic Research, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznań, Poland and [3]Institute for Cancer Research, Fox Chase Cancer Center, Philadelphia, PA 19111, USA.

*Corresponding author.

Key words: AIDS, ASV, HIV, integration, recombination

Received: **12 Oct 1995**
Revisions requested: **7 Nov 1995**
Revisions received: **22 Nov 1995**
Accepted: **27 Nov 1995**

Structure 15 January 1996, 4:89–96

Introduction

The retroviral integrase (IN) is a virus-encoded enzyme that catalyzes insertion of viral DNA into many sites on host DNA [1–3]. Because DNA integration is an essential step in the retroviral replication cycle, this enzyme is an attractive target for inhibition of human immunodeficiency virus (HIV), the causative agent of acquired immunodeficiency syndrome (AIDS). Work over the last few years has resulted in a general understanding of the enzymatic integration mechanism but more detailed analyses have been hampered by the lack of precise structural information. The situation changed recently when the crystal structures of the catalytic domains of both HIV-1 IN [4] and avian sarcoma virus (ASV) IN [5] became available; however, neither structure contained bound divalent cations, although such cofactors are necessary for the activity of these enzymes. Precise data on the interaction of these enzymes with these essential ligands are necessary for understanding the structural basis of the reaction mechanism, and for guiding rational drug design.

The integration reaction takes place in two steps. In the first step (processing) IN nicks the linear viral DNA, usually two base pairs from the 3′ end of each viral strand, adjacent to the highly conserved CA dinucleotide [6]. In the second step (joining) [7,8] there is a direct attack by the newly formed 3′-OH end of viral DNA on host DNA phosphates, resulting in covalent attachment [9]. The processing and joining reactions can be reproduced *in vitro* using short oligodeoxynucleotide duplexes as substrates [6–8]. These *in vitro* reactions require only IN and a divalent metal cofactor, Mn^{2+} or Mg^{2+}. Although the presence of the cofactor is essential, the placement and number of divalent cations bound in the active site of the integrase are not known.

Evolutionary comparisons of deduced amino-acid sequences of integrases have revealed a conserved motif (D,D(35)E) of three invariant carboxylate residues, two of which are separated by 35 amino acids D,D(35)E ([10,11] and references therein). This conservation extends through retroviral and retrotransposon integrases as well as some bacterial transposases, and suggests common mechanistic features for the active sites of these enzymes. Additional studies showed that these residues are essential for both the IN processing and joining reactions, indicating the presence of a single active site [11–14]. On the basis of the proposed mechanisms for the catalytic activity of the exonuclease domain of *E. coli* DNA polymerase I [15] and RNase H [16–18] we hypothesized that the conserved carboxylate residues of IN play a role in metal binding and catalysis [10,11]. This proposal has been supported by the crystal structure of the catalytic domain of ASV IN [5] which shows that the three conserved carboxylates are juxtaposed. We have also suggested that an enzyme–metal complex could stabilize a pentacoordinate phosphate intermediate during two successive nucleophilic attacks. The first would be on a viral DNA phosphate by a water molecule (in the processing reaction) and the second on a host DNA phosphate by the newly exposed 3′-OH at the end of viral DNA (in the joining reaction).

Here, we describe the structure of an isolated ASV IN catalytic domain complexed with metal ions. Although defective for the normal processing and joining activities, this isolated domain displays metal-dependent DNA cleavage-ligation activity ('disintegration'), as well as general nuclease activity. We find that Asp64 and Asp121 of the D,D(35)E motif coordinate a single Mn^{2+} or Mg^{2+} ion. This is the first direct evidence supporting the proposed role of the catalytic residues of IN. The placement of the ion is similar to that reported for metal complexes of RNases H of *E. coli* and HIV-1, enzymes which are structurally related to integrase as they share a similar folding motif that juxtaposes analogous conserved carboxylate residues in the active site [4,5,16,17,19–21].

We also perform a comparison between parts of the active site of ASV IN and the analogous regions in other members of this superfamily. Previously, we reported significant differences between the positions of analogous catalytic residues in the active sites of ASV and HIV-1 IN [5]. Here we show that these differences persist when the metal complex of ASV IN and uncomplexed HIV-1 IN are compared, and we suggest that the picture seen in the HIV-1 IN structure may be influenced by crystal packing and/or a tendency for disordering.

Results and discussion

Structure solution and description

Crystals of the catalytic domain of ASV IN (residues 52–207) were soaked in 10 mM $MnCl_2$ (2–3 times higher than assay conditions), as well as 20, 100, and 500 mM $MgCl_2$. Structures of the metal-soaked crystals were solved and refined at high resolution (Table 1). The catalytic domain of ASV IN consists of a five-stranded β sheet surrounded by five α helices, with the catalytic site located on the open surface of the protein (Fig. 1; see [5]). The electron-density map corresponding to the structure of the Mg^{2+} complex obtained at high metal concentration (MGH) was exceptionally clear (Fig. 2a) and showed a single Mg^{2+} ion bound between the carboxylates of Asp64 and Asp121 of the D,D(35)E motif. No indication of the presence of a second cation could be found in this map. Binding of Mg^{2+} at a lower concentration (20 mM, MGL) was much weaker, although the appearance of the electron-density map in this region bears closer resemblance to the density found in the presence of a high concentration of Mg^{2+} than to that found in the absence of metal. A putative metal/water cluster did not refine well, indicating that the level of substitution is considerably lower. A map obtained when the concentration of $MgCl_2$ was 100 mM is almost identical to that obtained at 500 mM, indicating full substitution of the metal-binding site (data not shown).

Table 1

Refinement statistics for metal complexes of ASV IN.

	MN*	MGH*	MGL*
Resolution range (Å)	8.0–2.05	8.0–1.70	8.0–2.20
No. of reflections			
Total	11953	19297	9523
2σ(F) cut-off	10294	17711	8529
Used for refinement	9450	16330	7624
Used for R-free	844	1381	635
No. of atoms			
Protein	1121	1128	1125
Water	141	173	124
Hepes	15	15	15
Metal	1	1	–
R-factor	0.130	0.150	0.138
R-free (8% reserved)	0.189	0.191	0.201
Completeness of data			
Measured reflections	99.9	99.8	99.3
Observed (2σ) reflections	86.0	91.6	86.1
Rms deviations from ideality			
Bond (Å)	0.013	0.013	0.013
Angle distances (Å)	0.047	0.037	0.041
Planes (Å)	0.016	0.015	0.017
Chiral volumes (Å^3)	0.165	0.136	0.149
PDB code	1VSF	1VSD	1VSE

*In 10mM $MnCl_2$ (MN), 500mM $MgCl_2$ (MGH), and 20mM $MgCl_2$ (MGL).

A single, similarly positioned metal ion (Fig. 2b) was detected after soaking the crystals in a 10 mM solution of $MnCl_2$, and the superb quality of the electron density map was comparable with that observed for the high $MgCl_2$ concentration. No direct interactions with the third important residue, Glu157, are observed with either

Figure 1

Chain tracing of the catalytic domain of ASV IN, showing the secondary structure elements and the location of the active site. (Figure prepared using the program RIBBONS [33].)

cation, although this residue has excellent definition in the electron-density maps. The peak height of the Mn^{2+} ion in the final $2F_o–F_c$ map was 6.5σ, indicating a very high level of confidence in its placement. As with Mg^{2+}, we detected only one Mn^{2+} ion bound to the protein. The higher affinity for the (larger) Mn^{2+} ion, demonstrated by its tight binding even at low concentration, is consistent with our previous observations that in the presence of Mn^{2+} ASV IN activity is 10–20 greater than when Mg^{2+} is present [6].

In addition to interacting with the two carboxylate oxygens, the metal ion interacts with four clearly seen water molecules, that complete its octahedral coordination (Fig. 2). The Mn^{2+}–O distances range from 2.0 Å to 2.4 Å (Table 2), showing very good agreement with the expected values derived from the Cambridge Crystallographic Database [22] even though they were not restrained in the refinement; Mg^{2+}–O distances are about 0.2 Å shorter (Table 2). The O–metal–O angles in the coordination octahedra are only slightly distorted, ranging from 85° to 95° in the Mg^{2+} complex, and from 80° to 102° in the Mn^{2+} complex. The coordinated water molecules make few hydrogen bonds with atoms outside the cluster. Wat253 makes clear hydrogen bonds with the carbonyl oxygen of Phe65, as well as with Wat349, and Wat254 makes a rather long hydrogen bond (3.15 Å) with Wat414.

Figure 2

Active site of the catalytic domain of ASV IN. **(a)** Stereoview of the electron-density map (generated using O [34]) for the Mg^{2+} complex (500 mM $MgCl_2$, see text). This $F_o–F_c$ map, contoured at 5σ level, was calculated at 1.8 Å resolution after refinement of a model which excluded the Mg^{2+} cation and its coordinated water molecules. The density corresponding to the cluster of an octahedrally-coordinated metal ion and four waters is exceedingly clear. **(b)** Stereoview of the active site of ASV IN generated using MOLSCRIPT [35]. Shown is part of the active site displaying the coordination of Mn^{2+} with four water molecules, as well as with the carboxylates of Asp121 and Asp64. The water molecule marked W324 is found in the same location in all ASV IN structures. The putative hydrogen bonds made by this molecule (red dashed lines), identified by an analysis of distances and angles, form a distorted tetrahedron (also including a bond to Nε2 of Gln153, not marked).

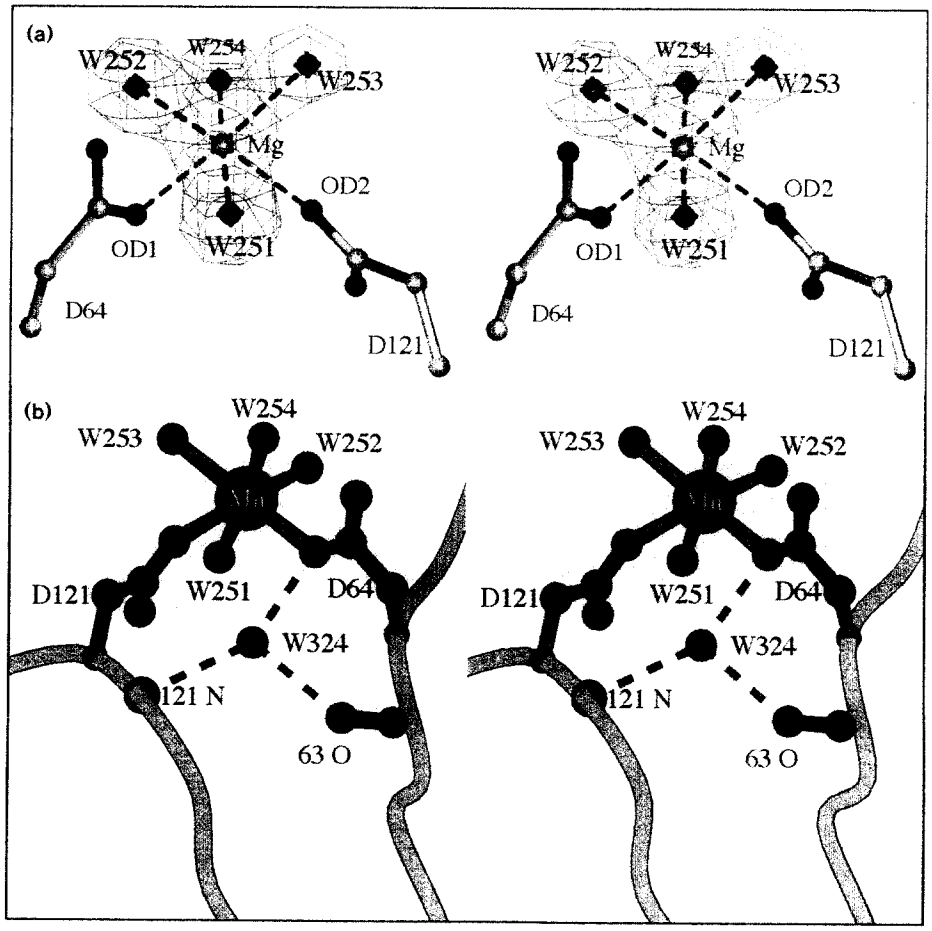


Table 2

Distances (Å) between the metal ions and the coordinating oxygens in the Mn and Mg complexes of ASV IN.

		MGH*	MN*
Asp64	OD1	2.15	2.27
Asp121	OD2	2.11	2.41
Wat251	O	1.98	2.00
Wat252	O	2.14	2.33
Wat253	O	2.14	2.40
Wat254	O	2.14	2.21

*In 500 mM $MgCl_2$ (MGH) and in 10 mM $MnCl_2$ (MN).

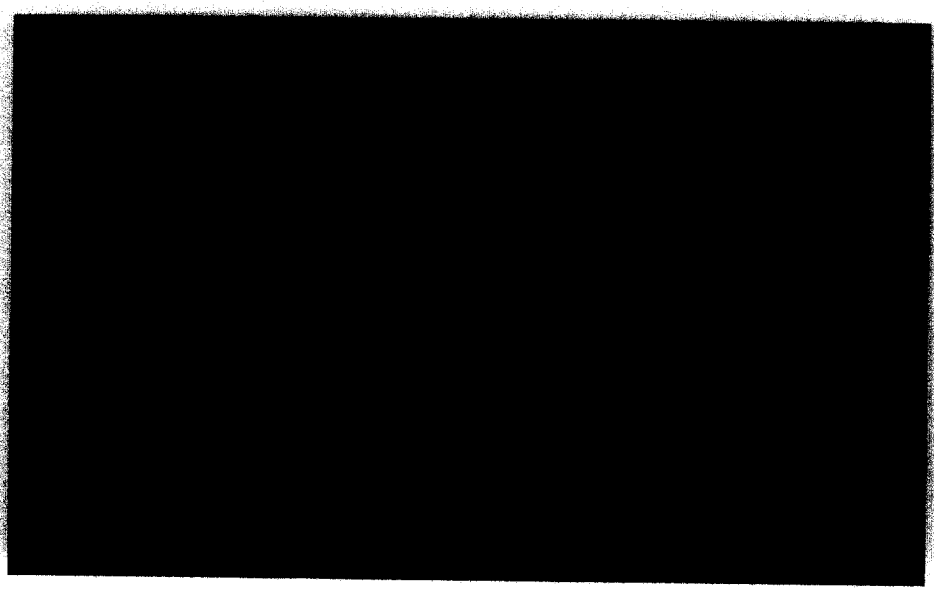
The atomic coordinates of the three crucial carboxylic acids are only slightly affected when the cation is bound (Fig. 3). There is practically no change in the position of the carboxylate of Asp64 (the maximum atomic movement is less than 0.4 Å), a slight rotation of the carboxylate of Asp121 leading to a movement of about 1 Å, and a more pronounced shift of the carboxylate of Glu157 in the Mn^{2+} structure, but not in the Mg^{2+} structure. The minimal deviation in carboxylate positioning upon metal binding is consistent with the observation that even single conservative substitutions in these residues drastically reduced activity of both ASV IN and HIV-1 IN [11]. Another potentially important residue, Gln153, is conserved in retroviral integrases and appears to stabilize the structure of the active site through water-mediated hydrogen bonds. The tightly bound Wat324, observed in the structures of the apoenzyme [5], is also present in an almost identical location in the presence of either cation. It is most likely to accept hydrogen bonds from the NE2 of Gln153 and N of Asp121, and donate them to O of Thr63 and OD1 of Asp64 (Fig. 2b). Other polar atoms are present in its vicinity, but it is not possible to postulate alternate networks of hydrogen bonds that involve sensible angles.

Comparison with related enzymes

Many enzymes active in the processing of nucleic acids, such as DNA polymerases, reverse transcriptases, or the nucleases discussed here, have active sites containing acidic residues and have a concomitant requirement for divalent cations for activity, but such cations have been reported only in a few structures. For example, only one structure of a reverse transcriptase, from Moloney murine leukemia virus (MMLV RT), with a single metal in its active site has been published [23], and none of the available structures of HIV-1 RT show bound metals. In addition, the structure of MMLV RT with bound metal, as well as the structure of a metal complex of *E. coli* RNase H, are only available at lower resolution than the corresponding uncomplexed structures. As our data show, the quality of the metal-containing ASV IN structures is as excellent as that of the metal-free enzyme. Because the maps are of exceptional quality, we are able to describe the active site with a high level of detail and accuracy.

We have compared directly the structure of the ASV IN active site with the active sites of the other members of the superfamily for which metal complexes have been described, namely both HIV-1 and *E. coli* RNases H, or inferred, as is that case for *E. coli* RuvC resolvase. As noticed by Yang and Steitz [21], the similarity of the cluster of acidic residues forming the active sites of the RNases H is striking. The placement and direction of the analogous carboxylates in RNases H and ASV IN are very similar (Fig. 4).

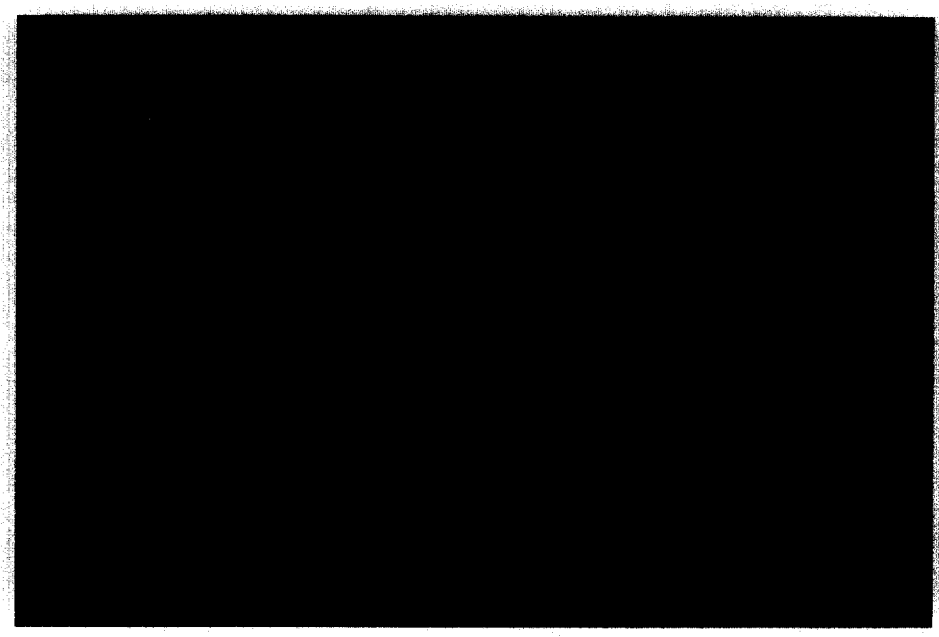
Figure 3



The residues found in the active site of integrases. They include three strictly conserved acidic residues (Asp64, Asp121, and Glu157), and Gln153, which may play a structural role. The coordinates displayed here were obtained by superimposing the Cα atoms of the entire proteins. The low temperature ASV IN structure (selenomethionine derivative; crystals grown from PEG/isopropanol) is in dark blue; room temperature ASV IN structure (crystals grown from ammonium sulfate), light blue; room temperature Mg^{2+} complex of ASV IN, green; Mn^{2+} complex of ASV IN, magenta. For HIV-1 IN (red, PDB code 1ITG), superimposed as above, only two of the four residues are visible, and Asp116 is positioned very differently from its equivalent in ASV IN, Asp121.

Figure 4

Comparison of the environment of divalent cations in the active site of ASV IN and *E. coli* and HIV-1 RNase H. ASV IN (structure MGH) is in red; *E. coli* RNase H, blue (PDB code 1RDD); HIV-1 RNase H, yellow (PDB code 1HRH). Two divalent cations are present in the latter structure (see text below). The presence of the cations is necessary for activity.



In agreement with our observations for ASV IN, the positions of the residues in the active site of *E. coli* RNase H [18,24,25] are influenced very little by the presence of the metal, with the Cα atom shifts being less than 0.4 Å and the carboxylate group shifts within 1.5 Å. Interestingly, the two residues that actually coordinate the Mg^{2+} ion move less than the other carboxylates, indicating that the part of the active site directly coordinating the metal ion has an invariant character. The situation is similar for MMLV RT, as the differences observed in the positions of the three critical aspartates in the active site are not larger than 0.4 Å when the structures with and without metal complexed are compared; however, in the case of MMLV RT the quality of the difference Fourier map is not sufficient to define the details of the metal coordination [23]. The active sites of MMLV RT and ASV IN, are also similar in that only two of the three carboxylates interact with a single Mn^{2+} cation present in the active site.

The active site of HIV-1 RNase H contains two metal ions [16] whereas that of *E. coli* RNase H contains only one [24]. The position of Mn^{2+} in ASV IN is very close to the second Mn^{2+} in HIV-1 RNase H (Fig. 4, [16]) and to the sole Mg^{2+} ion reported for *E. coli* RNase H [24]. Although no precise data on the location of a divalent cation are available for RuvC resolvase, an Mn^{2+}-binding site apparently exists between Asp7 (equivalent to Asp64) and Asp141 (equivalent to Glu157) [19]. For the two (quite divergent) RNase H enzymes, the maximum differences in the positions of the carboxylates do not exceed 1.5 Å, despite some disorder reported in the vicinity of the active site of isolated HIV-1 RNase H [16]. A similar picture emerges from a comparison of the active sites of MMLV RT [23] and HIV-1 RT [26], with root mean square (rms) deviations at the three active-site residues not exceeding 1.5 Å.

The number of bound divalent cations reported and their coordinating residues do differ between each RNase H. This has prompted some controversy concerning details of the catalytic mechanism of these enzymes. On the face of it, our observation of only a single metal bound to ASV IN may seem to support a single-divalent-cation mechanism, as proposed for *E. coli* RNase H [24], rather than the two-cation mechanism suggested for HIV-1 RNase H [16]. But the possibility that in the full-length IN, or in the presence of a nucleic acid substrate, a second cation might replace the water molecule bound between Asp64 and Glu157 cannot be ruled out at this point. As the carboxylate of Glu157 does not directly interact with the metal ion in the isolated catalytic domain, our results do not explain why even conservative substitutions of this residue lead to inactivation of the enzyme. We note that other enzymes discussed above also have one or more conserved and necessary carboxylates that have not been found to bind any divalent cations.

Differences in the active sites of ASV IN and HIV-1 IN

As we have noted previously, a comparison of the structurally aligned sequences of the catalytic domains of ASV IN (residues 52–207) and HIV-1 IN (residues 50–212) reveals a moderate 24% identity and an additional 8% similarity [5]. This level of similarity is comparable with that encountered in our previous investigations of the retroviral protease (PR), the structure of which was initially solved for ASV PR [27] and then used to model HIV-1 PR [28]. Comparisons of the crystal structures of those

enzymes [29] and of the recently solved PR structures from the feline immunodeficiency virus [30] and equine infectious anemia virus (Wlodawer *et al.*, unpublished data) have shown that although the sequence identity between the retroviral PRs also does not exceed 20–25%, their active sites are identical. When either apoenzymes or inhibitor complexes are compared the rms deviation between all atoms in the catalytic center does not exceed 0.6 Å [29,30]. As shown above, a similar situation has been observed for HIV-1 RT and MMLV RT, and for *E. coli* RNaseH and HIV-1 RNase H.

On the other hand, comparison of the current models of the active sites of ASV IN and HIV-1 IN derived from crystal structures of the isolated catalytic domains has presented a very different picture [5]. Our structural alignment of the crystallographic monomers of ASV IN and HIV-1 IN catalytic domains [5] showed that the conserved secondary structure elements could be superimposed with an rms deviation of 1.4 Å between 107 Cα pairs. The small degree of deviation is similar to that observed between the backbones of ASV and HIV-1 PR [29], FIV and HIV-1 PR [30], and MMLV and HIV-1 RT [23]. But significant differences in the structural models of ASV and HIV-1 IN were found when the conformations of the residues in the active site were compared. Although the β strand containing the Asp64 residue was observed to have a nearly identical main-chain trace in both models, the main chain that embeds ASV IN Asp121 and its HIV-1 counterpart, Asp116, was seen to follow different paths. In addition, as illustrated in Figure 4, the two catalytic aspartates were found to point in completely opposite directions in the two structures. The third catalytic residue, Glu152 in HIV-1 IN, could not be analyzed, as it is located in a region of 13 disordered amino acids. These differences do not reflect an absence of a metal cofactor in ASV IN, as the conformations of the active-site residues in this protein do not change significantly upon metal binding (see above). In the ASV IN structure, the metal docking site is apparently preformed. As a metal complex has not been reported for the HIV-1 IN catalytic domain, we do not know how metal binding might affect its structure; however, the similarities in the metal complexes of ASV IN and even distant members of this superfamily are likely to reflect shared mechanistic features of these enzymes.

In comparing all of our structures for ASV IN with the structure reported for the HIV-1 IN catalytic domain, we are struck by the fact that the paths of the main chain diverge most significantly in the vicinity of the active site. One possibility is that these differences may reflect different dynamic properties of the IN proteins, whereby the HIV-1 and ASV IN structures portray different conformers of the active site. We believe that the conformation of the ASV IN active site is relevant to activity because the required metals bind to a preformed site between Asp64 and Asp121, and the region of Glu157 is ordered. The significance of the conformation of HIV-1 IN is not yet apparent. Alternative possibilities are that simply the HIV-1 IN and ASV IN structures differ in the active-site region, that the termini chosen for the HIV-1 and ASV catalytic domain fragments influence their stability, or that the conformation of the active-site residues is influenced by crystal packing. Examination of the available coordinates of HIV-1 IN (PDB code 1ITG) shows that the active site is intimately involved in crystal contacts. Direct intermolecular hydrogen bonds are made in the vicinity of Asp116, between O of Thr115 and NE1 of Trp131, as well as between N of Gly118 and O of Trp131. In addition, OD2 of Asp116 is only 3.8 Å from the NZ atom of Lys136 of a neighboring molecule, forming a long salt bridge. Asp64, for which the C position is virtually identical to that of the corresponding residue in ASV IN, is further away from the crystal contacts. Crystal packing is unlikely to affect the active site of ASV IN, as the shortest distance between any of the three catalytic carboxylates and an atom belonging to another copy of the enzyme in the crystal lattice is 8.26 Å (between OD1 of Asp121 and NH1 of Arg168); however, no direct proof that the differences are due to crystal contacts will be available until different crystal forms of both enzymes are found.

It is also conceivable that, in order for the HIV-1 domain to adopt a conformation that would be structurally equivalent to that of ASV IN, binding of the DNA substrate or the presence of other domains of the protein may be required. We note that the full-length ASV IN is considerably more active than HIV-1 IN [11] and this could reflect the structural differences in the active sites described here.

Biological implications

Integrases (IN) are enzymes essential for retroviral replication. They are active as multimers (dimers or tetramers) and synchronize a two-step reaction resulting in the integration of the proviral DNA into the host genome. Both steps use a similar mechanism of nucleophilic attack on the phosphate backbone of DNA, and both depend on the same active site. Each monomer of integrase consists of three distinct domains, with the catalytic activity residing in the largest central catalytic domain (residues 52–207 in avian sarcoma virus integrase [ASV IN]). The isolated central domain possesses partial catalytic activities and, like the full-length enzyme, divalent cations, either Mg^{2+} or Mn^{2+}, are essential to its function. At the present time, crystal structures of the isolated catalytic domains of human immunodeficiency virus type 1 (HIV-1) and ASV IN in the absence of metals are available. In this first report describing the structure of an integrase in complexes with two different metal cations we present the details of the binding of both types of divalent cations to the active site of ASV

IN. Only one cation is coordinated in the active site of this isolated catalytic domain, docked into an apparently preformed metal ion binding site in the enzyme.

Structural comparisons of the active site of ASV IN with those of other members of this structurally related superfamily, such as RNase H and RuvC resolvase, show that there is a high degree of similarity in the general disposition of the catalytically-important aspartates (or glutamates), although the details of the binding of the divalent cations are different. The existence of a preformed metal-binding site in the catalytic center of this family of enzymes is strongly supported by the results reported here and by previous analyses of the structure of RNase H metal complexes. As the active site of the uncomplexed HIV-1 IN shows a conformation that is distinctly different from ASV IN (in either the absence or presence of bound metal) we postulate that the currently available HIV-1 IN structure may represent a different conformer, although other possible explanations are also discussed. Further investigation will be required to resolve this question and to determine the biological significance of these unexpected differences between the two retroviral integrase structures.

Materials and methods

Expression, purification, and crystallization of the catalytic core domain of ASV IN have been described previously [5]. The protein used for crystallizations was stored in a buffer containing 50 mM HEPES, 500 mM NaCl, 1% thiodiglycol, 0.1 mM EDTA and 40% glycerol. Just before crystallization experiments were set up, it was dialyzed to an analogous buffer containing 5% glycerol. Crystals were grown using the hanging drop vapor diffusion method from protein solutions at a concentration of 6–7.5 mg ml^{-1} mixed with precipitation buffer containing 100 mM HEPES, 10% isopropanol, 20% PEG 4000, in the ratio 7:4. $MnCl_2$ or $MgCl_2$ was dissolved in the well buffer to produce the soaking solutions. The crystals were soaked for three days in 10 mM $MnCl_2$, as well as 20, 100, and 500 mM $MgCl_2$.

Diffraction data were collected at room temperature on a MAR 300 mm image plate detector, using graphite-monochromated Cu Kα radiation. They were processed with DENZO, and scaled with SCALEPACK [31]. The low temperature structure of the Se-Met protein [5] was used as the initial model for the Mn^{2+} complex, which in turn was used as a starting model for the refinement of the structure of the $MgCl_2$-soaked crystals. Data with $F>2\sigma(F)$ were used in the refinement, and the low-resolution cut-off was 8 Å. The structure refinements were carried out in PROLSQ [32] modified to utilize free R-factor computation, with 8% of the data reserved for that purpose. These reserved reflections were separately chosen in a random manner for each data set. The final conventional R-factors were 0.130, 0.138 and 0.150 at the resolution of 2.05 Å, 2.2 Å and 1.7 Å, for $MnCl_2$, 20 mM $MgCl_2$ and 500 mM $MgCl_2$, respectively (Table 1). The coordinates have been deposited with the Brookhaven Protein Data Bank (see Table 1 for PDB codes).

Acknowledgements

We would like to acknowledge our Fox Chase colleague, B Mueller, G Kukolj, M Andrake and E Assante-Appiah who have provided encouragement and helpful advice throughout this project. Research was sponsored in part by the National Cancer Institute, DHHS, under contract NO1-CO-46000 with ABL. Other support included National Institutes of Health grants CA47486, CA06927, a grant for infectious disease research from Bristol-Myers Squibb Foundation, and an appropriation from the Commonwealth of Pennsylvania. The contents of this publication do not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government.

References

1. Katz, R.A. & Skalka, A.M. (1994). The retroviral enzymes. *Annu. Rev. Biochem.* **63**, 133–173.
2. Goff, S.P. (1992). Genetics of retroviral integration. *Annu. Rev. Genet.* **26**, 527–544.
3. Vink, C., Groeneger, O.A.M. & Plasterk, R.H. (1993). Identification of the catalytic and DNA-binding region of the human immunodeficiency virus type I integrase protein. *Nucleic Acids Res.* **21**, 1419–1425.
4. Dyda, F., Hickman, A.B., Jenkins, T.M., Engelman, A., Craigie, R. & Davies, D.R. (1994). Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. *Science* **266**, 1981–1986.
5. Bujacz, G., *et al.*, & Skalka, A.M. (1995). High resolution structure of the catalytic domain of the avian sarcoma virus integrase. *J. Mol. Biol.* **253**, 333–346.
6. Katzman, M., Katz, R.A., Skalka, A.M. & Leis, J. (1989). The avian retroviral integration protein cleaves the terminal sequences of linear viral DNA at the in vivo sites of integration. *J. Virol.* **63**, 5319–5327.
7. Craigie, R., Fujiwara, T. & Bushman, F. (1990). The IN protein of Moloney murine leukemia virus processes the viral DNA ends and accomplishes their integration *in vitro*. *Cell* **62**, 829–837.
8. Katz, R.A., Merkel, G., Kulkosky, J., Leis, J. & Skalka, A.M. (1990). The avian retroviral IN protein is both necessary and sufficient for integrative recombination *in vitro*. *Cell* **63**, 87–95.
9. Engelman, A., Mizuuchi, K. & Craigie, R. (1991). HIV-1 DNA integration: mechanism of viral DNA cleavage and DNA strand transfer. *Cell* **67**, 1211–1221.
10. Khan, E., Mack, J.P., Katz, R.A., Kulkosky, J. & Skalka, A.M. (1991). Retroviral integrase domains: DNA binding and the recognition of LTR sequences. *Nucleic Acids Res.* **19**, 851–860.
11. Kulkosky, J., Jones, K.S., Katz, R.A., Mack, J.P. & Skalka, A.M. (1992). Residues critical for retroviral integrative recombination in a region that is highly conserved among retroviral/retrotransposon integrases and bacterial insertion sequence transposases. *Mol. Cell. Biol.* **12**, 2331–2338.
12. Engelman, A. & Craigie, R. (1992). Identification of conserved amino acid residues critical for human immunodeficiency virus type 1 integrase function *in vitro*. *J. Virol.* **66**, 6361–6369.
13. van Gent, D.C., Groeneger, A.A.M.O. & Plasterk, R.H.A. (1992). Mutational analysis of the integrase protein of human immunodeficiency virus type 2. *Proc. Natl. Acad. Sci. USA* **89**, 9598–9602.
14. Drelich, M., Wilhem, R. & Mous, J. (1992). Identification of amino acid residues critical for endonuclease and integration activities of HIV-1 IN protein *in vitro*. *Virology* **188**, 459–468.
15. Beese, L.S. & Steitz, T.A. (1991). Structural basis for the 3′-5′ exonuclease activity of *Escherichia coli* DNA polymerase I: a two metal ion mechanism. *EMBO J.* **10**, 25–33.
16. Davies, J.F., II, Hostomska, Z., Hostomsky, Z., Jordan, S.R. & Matthews, D.A. (1991). Crystal structure of the ribonuclease H domain of HIV-1 reverse transcriptase. *Science* **252**, 88–95.
17. Yang, W., Hendrickson, W.A., Crouch, R.J. & Satow, Y. (1990). Structure of ribonuclease H phased at 2 Å resolution by MAD analysis of the selenomethionyl protein. *Science* **249**, 1398–1405.
18. Katayanagi, K., *et al.*, & Morikawa, K. (1992). Structural details of ribonuclease H from *Escherichia coli* as refined to an atomic resolution. *J. Mol. Biol.* **223**, 1029–1052.
19. Ariyoshi, M., Vassylyev, D.G., Iwasaki, H., Nakamura, H., Shinagawa, H. & Morikawa, K. (1994). Atomic structure of the RuvC resolvase: A Holliday junction-specific endonuclease from *E. coli*. *Cell* **78**, 1063–1072.
20. Rice, P. & Mizuuchi, K. (1995). Structure of the bacteriophage Mu transposase core: A common structural motif for DNA transposition and retroviral integration. *Cell* **82**, 209–220.
21. Yang, W. & Steitz, T.A. (1995). Recombining the structures of HIV integrase, RuvC and RNase H. *Structure* **3**, 131–134.
22. Allen, F.H., *et al.*, & Watson, D.G. (1991). The development of Version 3 and 4 of the Cambridge Structural Database System. *J. Chem. Inf. Comput. Sci.* **31**, 187–204.
23. Georgiadis, M.M., Jessen, S.M., Ogata, C.M., Telesnitsky, A., Goff, S.P. & Hendrickson, W.A. (1995). Mechanistic implications from the structure of a catalytic fragment of Moloney murine leukemia virus reverse transcriptase. *Structure* **3**, 879–892.

24. Katayanagi, K., Okumura, M. & Morikawa, K. (1993). Crystal structure of *Escherichia coli* RNase HI in complex with Mg^{2+} at 2.8 Å resolution: proof for single Mg(2+)-binding site. *Proteins* **17**, 337–346.
25. Katayanagi, K., *et al.*, & Morikawa, K. (1990). Three-dimensional structure of ribonuclease H from *E. coli*. *Nature* **347**, 306–309.
26. Unge, T., *et al.*, & Strandberg, B. (1994). 2.2 Å resolution structure of the amino-terminal half of HIV-1 reverse transcriptase (fingers and palm subdomains). *Structure* **2**, 953–961.
27. Miller, M., Jaskólski, M., Rao, J.K.M., Leis, J. & Wlodawer, A. (1989). Crystal structure of a retroviral protease proves relationship to aspartic protease family. *Nature* **337**, 576–579.
28. Weber, I.T., Miller, M., Jaskólski, M., Leis, J., Skalka, A.M. & Wlodawer, A. (1989). Molecular modeling of the HIV-1 protease and its substrate binding site. *Science* **243**, 928–931.
29. Weber, I.T. (1990). Evaluation of homology modeling of HIV protease. *Proteins* **7**, 172–184.
30. Wlodawer, A., *et al.*, & Dunn, B.M. (1995). Structure of an inhibitor complex of the proteinase from feline immunodeficiency virus. *Nat. Struct. Biol.* **2**, 480–488.
31. Otwinowski, Z. (1992). *An Oscillation Data Processing Suite* for *Macromolecular Crystallography*. Yale University, New Haven.
32. Hendrickson, W.A. (1985). Stereochemically restrained refinement of macromolecular structures. *Methods Enzymol.* **115**, 252–270.
33. Carson, M. (1991). *Ribbons 2.0 Manual*. University of Alabama at Birmingham, USA.
34. Jones, T.A. & Kjeldgaard, M. (1994). O – *The Manual*. Uppsala University, Uppsala.
35. Kraulis, P.J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* **24**, 946–950.